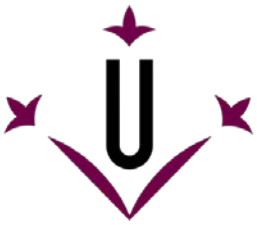


Universitat de Lleida

Document downloaded from:

http://hdl.handle.net/10459.1/63424

The final publication is available at:

https://doi.org/10.1111/age.12347

Copyright

(c) Stichting International Foundation for Animal Genetics, 2015

Expression profiling of the *GBP1* gene as a candidate gene for porcine reproductive and respiratory syndrome resistance

S. Gol, J. Estany, L.J. Fraile and R.N. Pena

Department of Animal Production, University of Lleida-Agrotecnio Center, 25198 Lleida, Spain.

Address for correspondence

R.N. Pena

Department of Animal Production, Campus ETSEA, University of Lleida-Agrotecnio Center, 25198 Lleida, Spain.

E-mail: romi.pena@prodan.udl.cat

Running head: pig GBP1 expression and PRRSV

Summary

A genomic region in the pig chromosome 4 has been previously associated to higher viremia levels and lower weight gain following PRRSV infection. The region includes the marker WUR1000125, a G>A polymorphism next to a putative polyadenylation site in the 3' untranslated region (3'UTR) of the interferon-induced guanylate-binding protein 1 (*GBP1*) gene. The protein encoded by *GBP1* is a negative regulator of T-cell responses. We show here that *GBP1* expression is lower in liver and tonsils of pigs carrying the WUR1000125-G allele, due to differential allele expression (allele A expression is 1.9-fold higher than allele G). We also show that the *GBP1* gene has two active polyadenylation signals 421-bp apart and that polyadenylation usage is dependent on the WUR1000125 genotype. The distal site is the most prevalently used in all samples but the presence of the A allele favours the generation of shorter transcripts from the proximal site. This is confirmed by a differential allele expression study in AG liver and tonsil samples. The interaction between WUR1000125 and other mutations identified in the 5' and 3'UTR regions of this gene should need to be studied. In conclusion, our study indicates that the WUR1000125 mutation is associated to changes in the expression of the negative T-cell regulator *GBP1* gene. However, the chromosome 4 locus for PRRSV viremia levels and weight gain contains a cluster of four other *GBP* genes that remain to be studied as candidate genes for this QTL.

Keywords genetic resistance, polyadenylation usage, porcine production, PRRSV, viral load

Introduction

The porcine reproduction and respiratory syndrome (PRRS) is one of the main clinical problems in pig production. Vaccines have proved unreliable to be a consistent approach to control the disease due to the high mutational rate of this virus. Consequently, it has been raised awareness that selection for resistant or tolerant pigs can be a solution to mitigate the negative impact of this virus on pig production. Recently, a major QTL at porcine chromosome SSC4 has been reported to affect both pig growth rate and response to American strains of PRRS virus (Boddicker *et al.* 2012). A group of 6 SNPs in perfect linkage disequilibrium captured 15.7% genetic variance for viremia levels and 11.2% for body weight gain after experimental infection. These polymorphisms are located in the interferon-induced guanylate-binding protein 1 (*GBP1*) gene, which has been associated to the control of the immune innate response to bacterial and viral infections in other species (Kim *et al.* 2011; Pan *et al.* 2012; Selleck *et al.* 2013). Among these 6 SNPs, WUR1000125 was selected as a tag SNP to evaluate the effect of alternate haplotypes. WUR1000125 is a G>A polymorphism which lay next to a putative polyadenylation site (AATAAA) in the 3' untranslated region (3'UTR) of *GBP1*. Mutations in the 3'UTR region can potentially affect transcript stability, thus influencing protein synthesis rate. Moreover, alternative usage of polyadenylation sites is a well-reporter regulator of protein expression, influencing mRNA stability, transport and translation, generally through the loss and gain of regulatory motifs, including microRNA-binding sites (Barrett *et al.* 2012; Sun *et al.* 2012). In the present work we have analysed whether the WUR1000125 mutation affects *GBP1* mRNA expression by analysing total expression levels, allele-specific expression, and polyadenylation site usage rate in liver and tonsils from pigs of different WUR1000125 genotypes.

Materials and Methods

DNA and RNA samples

Pig tissue samples were available in the lab at the time of this experiment. We used liver (n=42) and tonsil (n=13) samples from 42 Duroc pigs, collected upon slaughter and stored at -80ºC until analysis. Animals were all males from two batches of previous experiments developed under commercial conditions. Pigs in the same batch were from the same sex, and unit, and were slaughtered at the same age (100 days old, for batch one and 180 days old for batch 2). Within batch, pigs were selected from different litters to minimise parental relationship. Liver samples were collected from both batches while tonsils were collected only from pigs in batch 1. Genomic DNA was isolated from liver using standard protocols (Sambrook & Russell 2001). Total RNA was isolated with TRI-Reagent (Sigma-Aldrich) following the manufacturer's indications. Nucleic acid concentration and purity was assessed by spectrophotometry with Nanodrop-100 and the integrity tested by electrophoresis in agarose gels.

Retrotranscription

Prior to retrotranscription, 1 ug of total RNA was digested with Turbo DNA-free DNase (Ambion, LifeTechnologies) as indicated by the manufacturers in order to eliminate any traces of genomic DNA. The first-strand cDNA synthesis was performed with the RevertAid reverse transcriptase (Fermentas, ThermoFisher) in 20 ul reactions containing 1x buffer, 1mM dNTPs, 50 pmol random hexamers, 1 ul ribolock (Fermentas), 100U of enzyme and 1 μg of RNA. Reactions were incubated 10 min at 25ºC, 1h at 42ºC and 10 min at 70ºC.

GBP1 expression levels by real-time quantitative PCR (qPCR)

cDNA was diluted 1:10 in DEPC-treated H_2O prior to qPCR analysis. Primers (Supplementary Table S1) for *GBP1* and two reference genes, *YWHAZ* and *RPL32*, were designed with Primer3plus using the qPCR default parameters (Untergasser *et al.* 2007). For each gene, a standard curve was generated by amplifying serial dilutions of a control cDNA to check for linearity between initial template concentration and Ct values. Quantitative real-time PCR assays were carried out in triplicate in an ABI-7500 device (LifeTechnologies) in a final volume of 5 µl containing 1x Maxima SYBRgreen/ROX Master mix (Fermentas) and 200 nM of each primer. The following thermal profile was used for all reactions: 10 min at 95°C, 40 cycles of 15 s at 95°C and 1 min at 60°C, followed by a slow denaturation ramp from 60ºC to 95ºC to generate a dissociation curve to control the specificity of the amplified product. Cycle threshold (Ct) values were used as quantitation units. In order to quantify and normalise the expression data we used the ΔΔCt method (Yuan *et al.* 2006) using the geometric mean Ct value from the two reference genes and the *GBP1* Ct values.

WUR1000125 genotyping and allele-specific expression assays

The WUR1000125 was genotype with a custom allelic discrimination assay (LifeTechnologies). Primers and probes are given in Supplementary Table S1. We followed the manufacturer's protocol to set up the PCR reactions. Sequence Detection Systems software (SDS 2.0) was used to automatically collect and analyse the data and to generate the genotype calls. Allele-specific expression was analysed as in Lo *et al.* (2003). Briefly, genomic DNA from two homozygous pigs for WUR1000125, one with genotype AA and the other with genotype GG, were mixed at the following ratios: 8:1, 4:1, 2:1, 1:1,

1:2, 1:4, and 1:8 (AA:GG). Allelic discrimination assays were conducted as above and the Ct data was used to calculate, for each mixing ratio, the Ct(VIC)/Ct(FAM) ratio, where the VIC signal corresponds to the detection to the A allele and FAM to the detection of G. These data were used to generate a standard curve. The allelic discrimination assay was then run in heterozygous AG cDNA samples (liver= 9, from batch 1 and 2; tonsils=5, from batch 1) and the gene expression allele ratio was extrapolated by intercepting Ct(VIC)/Ct(FAM) ratio of each sample on the standard curve. Genomic DNA samples from the same heterozygous AG pigs were assayed in parallel as a control.

3′-rapid amplification of cDNA ends (3'RACE)

3'-end characterisation of the *GBP1* mRNA was carried out as follows from total RNA of tonsils (AA=4; AG=4) and liver (AA=4; AG=4; GG=4), selected at random within each genotype from pigs in batch 1 (tonsils) and batch 2 (liver). Retrotranscription was performed from total mRNA as above but using an anchored Oligo(dT) primer (Supplementary Table S1) that included an extended adaptor sequence (UAP – for universal amplification primer). Samples were incubated at 42°C for 1 h and the reaction was terminated at 70°C for 10 min. All primers used in this study are detailed in Supplementary Table S1 and Supplementary Figure S1. To amplify the 3' ends, PCR reactions were performed in 20 μl containing 0.4 μM of each primer (3'RACE_1/UAP) and 1 μl of cDNA. After an initial denaturation step at 94°C for 5 min, the reaction was performed for 1 cycle with 94ºC x 30 sec, 60ºC x 2 min and 72ºC x 2 min, followed by 30 cycles of 94ºC x 30 sec, 60ºC x 40 sec and 72°C x 90 sec. The final extension was carried out at 72°C for 10 min. A nested reaction was performed using 1 μl from the first reaction (1:100 dilution), with 0.4 μM of each primer (3'RACE_2/UAP) under the same cycling conditions. Amplified PCR products were subjected to 1.2% agarose gel electrophoresis

and visualised with ethidium bromide staining. Distinct PCR bands were excised from the gel, purified using QIAquick Gel Extraction Kit (QIAgen) and sequenced.

GBP1 polyadenylation usage ratio

A PCR was conducted with a FAM-labelled forward primer (GBP1_fam) and two reverse primers (Supplementary Table S1 and Supplementary Figure S2), one located just downstream of the first polyadenylation site and which included a string of seven Ts (GBP1-SHORT_R) and a second one just downstream the first primer, which should only be present in transcripts using the second polyadenylation site (GBP1-LONG_R). Samples included total RNA from liver (AA=4; AG=4; GG=4) and tonsils (AA=3; AG=3), selected at random within each genotype from pigs in batch 1 (tonsils) and batch 2 (liver). The PCR was performed with 0.3 ul of cDNA retrotranscribed with an anchored OligodT (50 pmol), diluted 1:10 in DEPC-treated H_2O using the same cycle conditions as described in the 3'RACE section. The expected sizes for the two PCR products were 725 and 745 bp, respectively. Given the small length differences, the two products are expected to be amplified with the same efficiency. After PCR, 1 ul of each reaction was mixed with 10 ul of HI-DI Formamide (LifeTechnologies), denatured at 95ºC x 5 min and incubated on ice x 2 min. The products were then solved in an ABI-3100 capillary electrophoresis system (LifeTechnologies). Peak identification and area under the curve were calculated with the DAx Data Acquisition and Data Analysis software.

Additionally, in the heterozygous samples (liver=4; tonsils=3) the transcripts resulting from the alternative polyadenylation sites were amplified separately with primers qGBP1_F/GBP1-SHORT_R and qGBP1_F/GBP1-LONG_R. The A and G allelic contribution to each of the transcripts was measured using an allelic discrimination assay from PCR template diluted 1:100, as explained in the allelic-specific expression section.

GBP1 promoter and full 3'UTR amplification and sequencing

Primers were designed with Primer3plus (Untergasser *et al.* 2007) to amplify and sequence the *GBP1* proximal promoter (up to 1,100 bp upstream the +ATG) and the full exon 11, which includes the STOP codon and the 3'UTR sequence (Supplementary Table S1). PCR reactions were carried out in a Veriti thermocycler (LifeTechnologies) in a volume of 25 µl containing 1x buffer, 200 µM dNTP mix, 2.0 mM $MgCl_2$, 400 nM of each primer, 1U of Taq polymerase (BIOTOOLS) and 60 ng of genomic DNA. The thermal profile was as follows: initial denaturing step for 5 min at 95°C and then 35 cycles of 20 s at 95ºC, 30 s at 60°C and 1.30 min 72ºC, finishing with 5 min at 72ºC. PCR products were solved by electrophoresis in a 1.2 % agarose gel and positive bands were purified with ExoSAPit enzyme kit (USB) and sequenced with the BigDye Terminator Sequencing kit v3.1 (LifeTechnologies). Sequences obtained were edited using the Sequencing Analysis software (LifeTechnologies) and aligned with the ClustalW program (Chenna *et al.* 2003).

Prediction of potential transcription factor-binding sites was performed with the TRANSFAC 8.3 database with the PROMO v2.0.3 tool (Messeguer *et al.* 2002) and the 2014 JASPAR-CORE repository (Mathelier *et al.* 2014). The 3'UTR region was scanned for RNA structural and regulatory motifs using RegRNA (Huang *et al.* 2006).

Statistical Analysis

Variation of *GBP1* expression was analysed in each tissue separately with a model including the batch and the WUR1000125 genotype. The least squares means of the genotypes were separated using the Tukey test. Allele-specific expression in AG samples was analysed within tissue as above in a model that include the batch, on the logit

transformation of the data, as these were closed percentages (Ros-Freixedes & Estany 2014). Comparison of polyadenylation usage by genotype, and of allele contribution per transcript, was performed within tissue using a *t*-test based on the logit transformation of the data. All the analyses were performed with JMP Pro 11 (SAS Inst. Inc., Cary, NC) and differences were considered significant at $p<0.05$.

Results

The genotype of the WUR1000125 mutation is associated to *GBP1* expression

GBP1 expression levels were measured in liver and tonsils of pigs with genotypes AA, AG and GG for the WUR1000125 SNP marker, which is located at the 3'UTR of the *GBP1* gene. Expression of *GBP1* was lower in the liver of pigs with the GG genotype as compared to AA ($p<0.01$) pigs (Figure 1). Expression of heterozygous AG liver samples were intermediate to the alternative homozygotes. In a subset of these animals, for which tonsils were available, we also observed higher expression of *GBP1* in AA pigs than in AG animals ($p<0.05$). No GG pig could be analysed for this tissue.

The expression of the WUR1000125 marker is allele-specific

Differential allele expression was assessed in liver and tonsils of heterozygous AG animals (Figure 2). In both tissues, allele A was expressed at higher levels (about 1.9-fold) than allele G ($p<0.001$). Presence of A and G alleles in controls of genomic DNA from AG pigs gave approximately a 50% ratio, as expected.

The *GBP1* gene has two active polyadenylation signals

Amplification of the 3' end of the *GBP1* mRNA indicated the presence of two main transcripts differing approximately 400 nt (Supplementary Figure S2). The two transcripts were detected in all samples analysed, which included liver from AA, AG and GG pigs and tonsils from AA and AG animals. PCR bands corresponding to the short and long transcript were subsequently sequenced to investigate the differences in size. Sequence alignment and analysis indicated that the two transcripts differed in length due to the alternative use of two active polyadenylation sites 421-bp apart. The full 3'UTR of the two transcripts were 603- and 1029-nt long. The A and G alleles were identified in the sequences of both long and short transcripts.

Transcription termination and start of polyadenylation took place 17 and 16 nt after the proximal and distal polyadenylation sites, respectively (Supplementary Figure S2).

The analysis of the sequence also revealed the presence of other polymorphisms in phase to the WUR1000125 in this fragment of the 3'UTR of the gene (Supplementary Figure S2) which are explained in more detailed below.

Polyadenylation usage is dependent on the WUR1000125 genotype

The WUR1000125 polymorphism lays 1-bp upstream of the first polyadenylation site. We next questioned whether the WUR1000125 polymorphism affected the rate of polyadenylation site usage. In order to test this, we carried out a modification of the 3'RACE protocol with a common FAM-labelled forward primer in exon 10 and two reverse primers in exon 11 that were specific for the short or the long transcripts (Supplementary Figure S1). After amplification, fragments were solved by capillary electrophoresis and quantified as FAM fluorescent units (Figure 3). The distal polyadenylation site exhibited the highest usage rate in all samples analysed. On the other hand, the prevalence of the short transcript generated from the proximal polyadenylation

site differed across genotypes ($p<0.05$). In AA pigs, short transcripts represented a third of all the *GBP1* mRNAs, both in liver and tonsils. The use of this proximal site was lower in AG pigs, representing 13% and 18% of total *GBP1* mRNAs in liver and tonsils, respectively. The lowest usage rate of the proximal polyadenylation site was observed in the liver of GG pigs resulting in approximately 10-fold higher expression of long versus short transcripts (Figure 3).

To further investigate this matter, the relative contribution of the A and G alleles to the total amount of short and long transcripts was investigated in liver and tonsil samples from AG pigs (Figure 4). Regarding the long transcripts, the ratio of A:G expression was similar to the allele-specific expression levels measured on total *GBP1* mRNA (Figure 2). The A allele contributed to ~60% of the transcripts in all samples. In contrast, the short transcripts had a higher representation of the A allele ($p<0.05$ in liver; $p<0.10$ in tonsils), particularly in liver where it accounted for 74% of the mRNA polyadenylated at the proximal site (Figure 4).

Description of other polymorphisms in the *GBP1* regulatory regions

The expression of *GBP1* is induced by interferon (IFN)-α/β and IFN-γ, as well as tumour necrosis factor (TNF)-α and interleukin-1 (Naschberger *et al.* 2004). In the human gene, several cytokine-response elements have been described in the proximal promoter region as well as in the 5'UTR sequence in exon 1. These include GAS (γ-IF activation sites), ISRE (IFN-α stimulated response element) and an NFκB-binding motif. Additional *in silico* analysis of potential transcription-factor binding sites identified two putative overlapping sites for interferon regulatory factor (IRF-1) and NFκB (Supplementary Figure S3). *GBP1* is a TATA-less promoter and, therefore, has several transcription start sites. We

sequenced 1,100 bp upstream of the ATG signal, which included the whole exon 1 and approximately 300-400 bp of promoter, in samples of pigs with AA and GG genotype for the WUR1000125 polymorphism. We identified a total of seven polymorphisms (six SNPs and a 1-bp INDEL) (Supplementary Table S2 and Supplementary Figure S3). None of the mutations changed the cytokine response elements described in this region but a SNP mutation at -631 bp co-localised with the putative overlapping IRF-1/NFκB regulatory elements (Supplemetary Table S3).

We next sequenced the entire 3'UTR region to describe any additional mutations that might affect the stability of this region. We identified 10 SNP polymorphisms (including the WUR1000125 marker). We searched for 3'UTR structural and regulatory elements co-localising with these mutations. The *GBP1* 3'UTR contains four interferon-response elements, three of which are common to the short and long transcripts (Supplementary Table S4). Three of the mutations identified lay in two of these regulatory elements (Supplementary Figure S4). In our sequencing data, the three mutations were fully linked to the WUR1000125 polymorphism.

Discussion

Several studies have reported differences in susceptibility to PRRSV infection and disease development between pig breeds (Reiner *et al.* 2010; Pena *et al.* 2013), but also within lines (Lewis *et al.* 2009). The genetic component of this has been confirmed by several authors (Lewis *et al.* 2009; Serao *et al.* 2014). Functional candidate genes for PRRSV susceptibility have been listed through global transcriptomics (Xiao *et al.* 2010; Arceo *et al.* 2012; Jiang *et al.* 2013) but only a few of them have been looked into in some more detail (Ren *et al.* 2012; Wang *et al.* 2012a; Wang *et al.* 2012b). The first genomic region

associated to response to PRRSV infection was reported in 2012 (Boddicker *et al.* 2012). Working with American commercial pigs, the authors identified a region in chromosome 4 with six markers in perfect linkage associated to PRRSV viremia levels and weight gain. One of these markers, WUR1000125, was selected as a tag-SNP to further characterize the influence of this region on viremia profiles of experimentally-infected pigs. The WUR1000125 marker (SNPdb accession number rs80800372) is a G>A SNP polymorphism that lays in the 3'UTR of the *GBP1* gene. Allele A, the unfavourable allele, is associated to higher PRRSV viremia levels and lower weight gain following infection (Boddicker *et al.* 2012). Conversely, the favourable G allele promotes lower plasma PRRSV titre levels and favours weight gain following a PRRSV challenge. In a previous study we have shown that the WUR1000125 marker segregates in European pig lines at allele frequencies similar to those described in American populations (Pena *et al.* 2013).

The *GBP1* gene encodes for an interferon-induced guanylate-binding protein belonging to the dynamin superfamily of large GTPases. This protein is an important player in cell-autonomous immunity (MacMicking 2012), displaying antiviral, antimicrobial and antiparasitic activity (Kim *et al.* 2011; Selleck *et al.* 2013; Zhu *et al.* 2013). Although interaction with several viral and microbial proteins have been postulated, the molecular mechanisms of action of GBP1 remained unknown until recent reports linking the action of this protein with the cytoskeleton remodelling that takes place in interferon-activated cells (Ostler *et al.* 2014). Moreover, GBP1 has been recently recognised as a negative regulator to T-cell activation (Forster *et al.* 2014), interfering with the early stage of T-cell receptor signalling through interaction with structural proteins.

Given all the above, it is likely that the haplotype described by Boddicker *et al.* (2012) might affect the functionality of this gene. Using WUR1000125 as a tag-SNP, we show here that pigs carrying the favourable G allele express less *GBP1* in liver and tonsils. The

lower mRNA levels are due to allele-specific differences in expression, with the A allele exceeding by approximately 1.9-fold the expression of the G allele. The differences of expression between alleles are comparable to the differences in expression between genotypes. Overall, the lower expression of the favourable G allele agrees well with the recent finding that GBP1 assists in tuning down T-cell responses (Forster *et al.* 2014). Genotypes associated to lower *GBP1* expression are expected to exhibit more effective T-cell responses. It is well-reported that a PRRSV-specific T lymphocyte IFN-γ response does not develop until at least 2 weeks after infection (Cecere *et al.* 2012). This cellular immune response is associated with an efficient immune response against this virus (Mateu & Diaz 2008). This fact highlights the relevance of efficient T-cell responses regarding the outcome of PRRSV infection.

The length of the 3′UTR is a major determinant in mRNA expression (Barrett *et al.* 2012). In general, longer 3′UTRs correlate with a relatively lower expression level, as longer 3′UTRs are more likely to possess miRNA binding sites and AU-rich elements that have the potential to promote mRNA decay and inhibit translation (Barrett *et al.* 2012). We show here that, in pigs, the *GBP1* gene has two active polyadenylation signals. Alternative usage of polyadenylation sites is one of the mechanisms leading to changes in the 3'UTR length of mRNA transcripts. The WUR1000125 mutation is next to the more proximal polyadenylation site, therefore we have tested if WUR1000125 could be causal to the drop of expression by affecting the usage rate of the two polyadenylation signals. The usage of both signals was confirmed by 3'RACE characterisation of *GBP1* in AA, AG and GG pigs. The three genotypes favoured the usage of the distal site both in liver and tonsils. Polyadenylation from the proximal signal was promoted mainly by the A allele, increasing the proportion of the shorter, potentially more-stable, transcripts by 4-fold in AA (36% of short transcripts) with respect to GG liver samples (~9%). The distribution of A and G

alleles among the longer transcripts paralleled the allele-specific expression results observed in the total *GBP1* mRNA. However, and in agreement with the polyadenylation usage results, the shorter transcripts had a larger representation of A alleles both in liver and tonsils of heterozygous pigs. These shorter transcripts have potentially less capacity to be regulated by trans-factors in response to internal cues, such as changes in the immune state of the pig. Conversely, the G allele tends to accumulate in longer transcripts originated from the second polyadenylation site. The potentially lower stability of these mRNAs agrees with the lower *GBP1* expression levels in pigs carrying the favourable G allele and is consistent with a model promoting the fine regulation of the GBP1 protein production.

Taken together, these results strongly indicate that the WUR1000125 mutation change the processability of the proximal polyadenylation signal of the *GBP1* gene. This can potentially be the cause of the differences in expression observed between genotypes, as the length of the 3'UTR influences the stability and the translation rate of the transcripts (Sun *et al.* 2012). Mutations in the gene promoter can also be responsible for differences in transcription levels. The WUR1000125 is in total linkage disequilibrium with other markers in the SSC4 chromosome (Boddicker *et al.* 2012). Our analysis of the *GBP1* gene has identified 16 other polymorphisms in the 5'UTR and 3'UTR of this gene, most of which are also linked to this tag-SNP. Although none of the three functional IFN and NFκB response elements known in the promoter are affected by these mutations, a putative IFN regulatory element overlaps a SNP mutation at -631 bp from the start ATG codon. Moreover, three mutations in the 3'UTR of the *GBP1* gene co-localise with two other IFNγ-response elements. The potential effect of these mutations on the expression of *GBP1* cannot be overlooked and their possible interaction with the WUR1000125 marker should be explored in more detail.

In conclusion, our study indicates that the QTL for PRRSV viremia levels and weight gain described by Boddicker *et al.* (2012) is associated to changes in the expression of the negative T-cell regulator *GBP1* gene. However, although the causal mutation responsible for this remains unknown, the results reported call attention to the transcriptional regulation of *GBP1*, a negative regulator of T-cell responses. Nevertheless, in the pig genome, *GBP1* is within a cluster of five *GBP* genes (*GBP1*, *2*, *4*, *5* and *6*). Further study of the other *GBP* genes might be of interest to help positioning the causal mutation. Moreover, it would be interesting to measure T-cell related cytokines in animals of AA and GG WUR1000125 genotype, to confirm the involvement of this genomic region with T-cell activation and response. A further question that remains unsolved is how the function of this gene family relates to the enhanced growth rate reported in PRRSV-challenged pigs (Boddicker *et al.* 2012). Our own preliminary data (Pena *et al.* 2013) suggests that the relationship of *GBP1* with growth rate depends on the epidemic phase of PRRS. While allele A is positively associated to growth in non-epidemic phases, allele G is, in contrast, enhancing weight gain during the epidemic phase or in challenged pigs. This hypothesis is further supported by the fact that the A allele, unfavourable for PRRSV viremia levels, is the most frequent in commercial populations selected for increased growth rate (Boddicker *et al.* 2012; Pena *et al.* 2013), with allelic frequencies in the range of 0.8-0.9. The contribution of this gene to growth gain in non-epidemic phases shall need to be assessed in the future.

Acknowledgements

S. Gol is recipient of a PhD scholarship from the Spanish Ministry of Education (BES-2014-FPU13/04975).

References

Arceo M.E., Ernst C.W., Lunney J.K., Choi I., Raney N.E., Huang T., Tuggle C.K., Rowland R.R. & Steibel J.P. (2012) Characterizing differential individual response to porcine reproductive and respiratory syndrome virus infection through statistical and functional analysis of gene expression. *Front Genet* **3**, 321.

Barrett L.W., Fletcher S. & Wilton S.D. (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* **69**, 3613-34.

Boddicker N., Waide E.H., Rowland R.R., Lunney J.K., Garrick D.J., Reecy J.M. & Dekkers J.C. (2012) Evidence for a major QTL associated with host response to porcine reproductive and respiratory syndrome virus challenge. *J Anim Sci* **90**, 1733-46.

Cecere T.E., Todd S.M. & Leroith T. (2012) Regulatory T cells in arterivirus and coronavirus infections: do they protect against disease or enhance it? *Viruses* **4**, 833-46.

Chenna R., Sugawara H., Koike T., Lopez R., Gibson T.J., Higgins D.G. & Thompson J.D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* **31**, 3497-500.

Forster F., Paster W., Supper V., Schatzlmaier P., Sunzenauer S., Ostler N., Saliba A., Eckerstorfer P., Britzen-Laurent N., Schutz G., Schmid J.A., Zlabinger G.J., Naschberger E., Sturzl M. & Stockinger H. (2014) Guanylate binding protein 1-mediated interaction of T cell antigen receptor signaling with the cytoskeleton. *J Immunol* **192**, 771-81.

Huang H.Y., Chien C.H., Jen K.H. & Huang H.D. (2006) RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res* **34**, W429-34.

Jiang C., Xing F., Xing J., Jiang Y. & Zhou E. (2013) Different expression patterns of PRRSV mediator genes in the lung tissues of PRRSV resistant and susceptible pigs. *Dev Comp Immunol* **39**, 127-31.

Kim B.H., Shenoy A.R., Kumar P., Das R., Tiwari S. & MacMicking J.D. (2011) A family of IFN-gamma-inducible 65-kD GTPases protects against bacterial infection. *Science* **332**, 717-21.

Lewis C.R., Torremorell M., Galina-Pantoja L. & Bishop S.C. (2009) Genetic parameters for performance traits in commercial sows estimated before and after an outbreak of porcine reproductive and respiratory syndrome. *J Anim Sci* **87**, 876-84.

Lo H.S., Wang Z., Hu Y., Yang H.H., Gere S., Buetow K.H. & Lee M.P. (2003) Allelic variation in gene expression is common in the human genome. *Genome Res* **13**, 1855-62.

MacMicking J.D. (2012) Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nat Rev Immunol* **12**, 367-82.

Mateu E. & Diaz I. (2008) The challenge of PRRS immunology. *Vet J* **177**, 345-51.

Mathelier A., Zhao X., Zhang A.W., Parcy F., Worsley-Hunt R., Arenillas D.J., Buchman S., Chen C.Y., Chou A., Ienasescu H., Lim J., Shyr C., Tan G., Zhou M., Lenhard B., Sandelin A. & Wasserman W.W. (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* **42**, D142-7.

Messeguer X., Escudero R., Farre D., Nunez O., Martinez J. & Alba M.M. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18**, 333-4.

Naschberger E., Werner T., Vicente A.B., Guenzi E., Topolt K., Leubert R., Lubeseder-Martellato C., Nelson P.J. & Sturzl M. (2004) Nuclear factor-kappaB motif and interferon-alpha-stimulated response element co-operate in the activation of guanylate-binding protein-1 expression by inflammatory cytokines in endothelial cells. *Biochem J* **379**, 409-20.

Ostler N., Britzen-Laurent N., Liebl A., Naschberger E., Lochnit G., Ostler M., Forster F., Kunzelmann P., Ince S., Supper V., Praefcke G.J., Schubert D.W., Stockinger H., Herrmann C. & Sturzl M. (2014) Gamma interferon-induced guanylate binding protein 1 is a novel actin cytoskeleton remodeling factor. *Mol Cell Biol* **34**, 196-209.

Pan W., Zuo X., Feng T., Shi X. & Dai J. (2012) Guanylate-binding protein 1 participates in cellular antiviral response to dengue virus. *Virol J* **9**, 292.

Pena R., Estany J. & Fraile L. (2013) Association between genetic resistance prrs virus and pig growth performance of commercial lines. In: *XV National Animal Production Meeting* (eds. by Calvo J, Casasús I, Joy M, Álvarez J, Varona L, Panea B, Calvete C & Balcells J), pp. 583-6. Asociación Interprofesional para el Desarrollo Agrario, Zaragoza (Spain).

Reiner G., Willems H., Pesch S. & Ohlinger V.F. (2010) Variation in resistance to the porcine reproductive and respiratory syndrome virus (PRRSV) in Pietrain and Miniature pigs. *J Anim Breed Genet* **127**, 100-6.

Ren Y.W., Zhang Y.Y., Affara N.A., Sargent C.A., Yang L.G., Zhao J.L., Fang L.R., Wu J.J., Fang R., Tong Q., Xiao J., Li J.L., Jiang Y.B., Chen H.C. & Zhang S.J. (2012) The polymorphism analysis of CD169 and CD163 related with the risk of porcine reproductive and respiratory syndrome virus (PRRSV) infection. *Mol Biol Rep* **39**, 9903-9.

Ros-Freixedes R. & Estany J. (2014) On the Compositional Analysis of Fatty Acids in Pork. *Journal of Agricultural Biological and Environmental Statistics* **19**, 136-55.

Sambrook J. & Russell D. (2001) *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Selleck E.M., Fentress S.J., Beatty W.L., Degrandi D., Pfeffer K., Virgin H.W.t., Macmicking J.D. & Sibley L.D. (2013) Guanylate-binding protein 1 (Gbp1) contributes to cell-autonomous immunity against Toxoplasma gondii. *PLoS Pathog* **9**, e1003320.

Serao N.V., Matika O., Kemp R.A., Harding J.C., Bishop S.C., Plastow G.S. & Dekkers J.C. (2014) Genetic analysis of reproductive traits and antibody response in a PRRS outbreak herd. *J Anim Sci* **92**, 2905-21.

Sun Y., Fu Y., Li Y. & Xu A. (2012) Genome-wide alternative polyadenylation in animals: insights from high-throughput technologies. *J Mol Cell Biol* **4**, 352-61.

Untergasser A., Nijveen H., Rao X., Bisseling T., Geurts R. & Leunissen J.A. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71-4.

Wang S.J., Liu W.J., Sargent C.A., Zhao S.H., Liu H.B., Liu X.D., Wang C., Hua G.H., Yang L.G., Affara N.A. & Zhang S.J. (2012a) Effects of the polymorphisms of Mx1, BAT2 and CXCL12 genes on immunological traits in pigs. *Mol Biol Rep* **39**, 2417-27.

Wang S.J., Liu W.J., Yang L.G., Sargent C.A., Liu H.B., Wang C., Liu X.D., Zhao S.H., Affara N.A., Liang A.X. & Zhang S.J. (2012b) Effects of FUT1 gene mutation on resistance to infectious disease. *Mol Biol Rep* **39**, 2805-10.

Xiao S., Jia J., Mo D., Wang Q., Qin L., He Z., Zhao X., Huang Y., Li A., Yu J., Niu Y., Liu X. & Chen Y. (2010) Understanding PRRSV infection in porcine lung based on genome-wide transcriptome response identified by deep sequencing. *PLoS ONE* **5**, e11377.

467 Yuan J.S., Reed A., Chen F. & Stewart C.N., Jr. (2006) Statistical analysis of real-time PCR data. *BMC*
468 *Bioinformatics* **7**, 85.

469 Zhu Z., Shi Z., Yan W., Wei J., Shao D., Deng X., Wang S., Li B., Tong G. & Ma Z. (2013)
470 Nonstructural protein 1 of influenza A virus interacts with human guanylate-binding
471 protein 1 to antagonize antiviral activity. *PLoS ONE* **8**, e55920.

472

473 **Table 1.** Pig tissue samples used in the experiment, distributed by WUR1000125 genotype

474

		WUR1000125 genotype		
Tissue	**n**	**AA**	**AG**	**GG**
Liver	42	20	16	6
Tonsils	13	8	5	-

475

476

Figure 1. LS Means of *GBP1* gene expression in liver and tonsils of Duroc pigs, by WUR1000125 genotype. Error bars represent standard errors. Within tissue, different letters indicate significant differences ($p<0.05$).

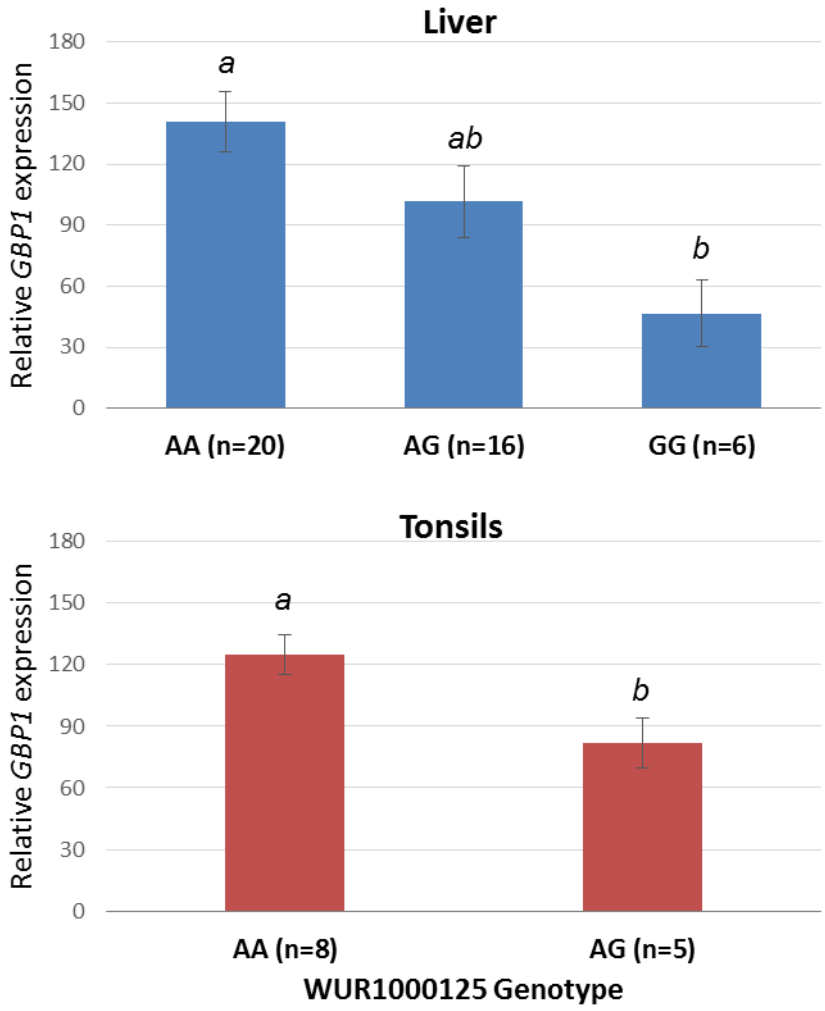


Figure 2: Allele-specific expression at the WUR1000125 polymorphism of the *GBP1* gene measured in total RNA from liver and tonsils of heterozygous pigs and in genomic DNA as a control. Error bars represent the standard error of each mean. Within tissue, different letters indicate significant differences ($p<0.05$).

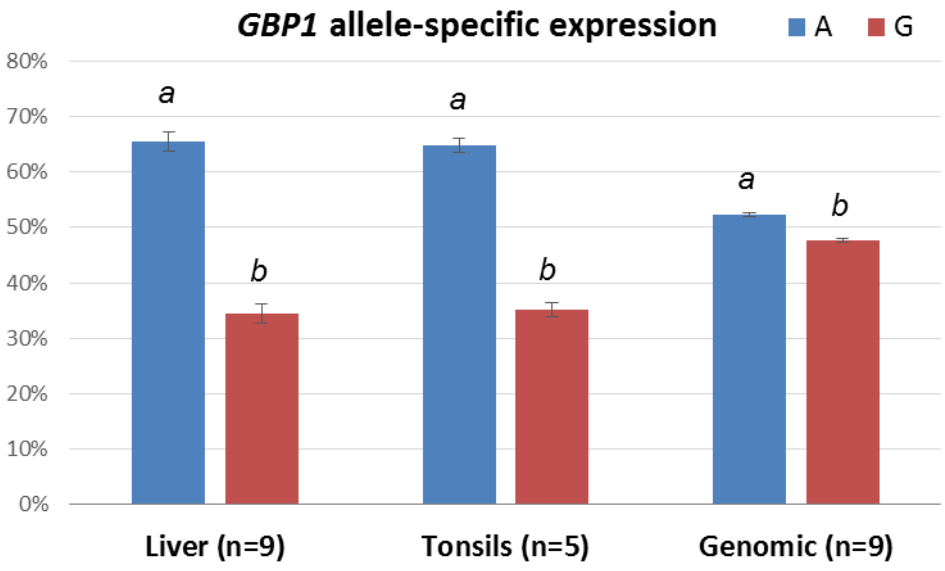


Figure 3. (A) Production of short and long transcripts by alternative use of proximal (pA1) and distal (pA2) polyadenylation sites in tonsil and liver samples from pigs with different genotype on the WUR1000125 marker of the *GBP1* gene. Error bars represent the standard error of each mean. Within tissue and transcript, different letters indicate significant differences between genotypes ($p<0.05$). Quantification of short and long transcripts after capillary electrophoresis of FAM-labelled 3'UTR PCR products in an AA tonsil (B) and a GG liver (C) sample.

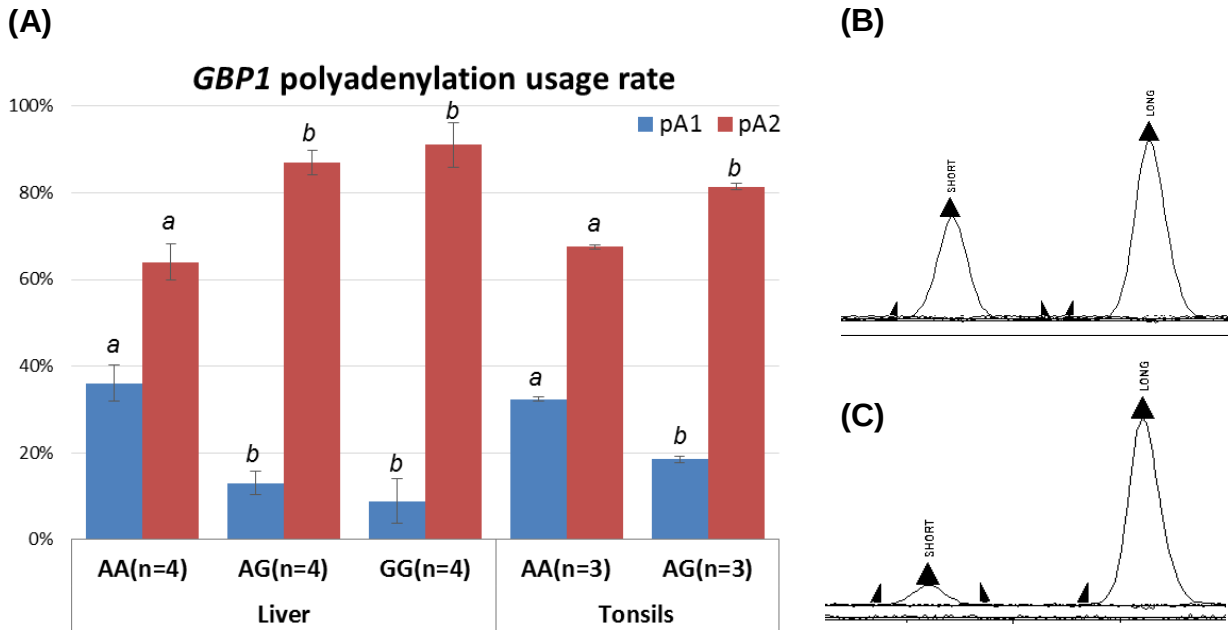
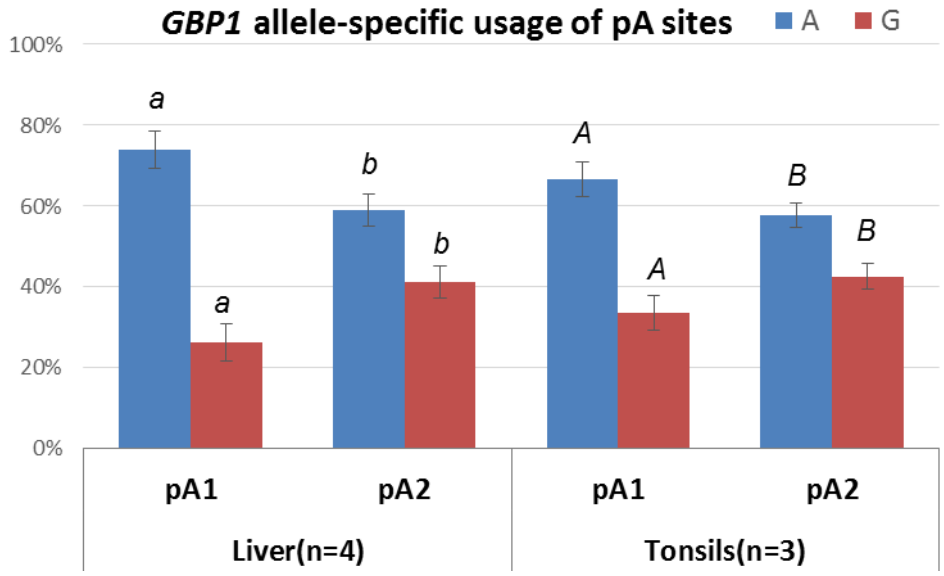


Figure 4. Contribution of WUR1000125 alleles to the short and long transcripts of *GBP1*, generated by alternative use of proximal (pA1) and distal (pA2) polyadenylation signals, in liver and tonsil heterozygous AG pigs. Data represent percentage of allele contribution to each transcript. Within each tissue and allele, different letters indicate significant differences in transcript distribution (lowercase, $p<0.05$; uppercase, $p<0.1$). Error bars represent the standard error of each mean.



511 **Supplementary Table 1.** Primers and probes used in this study.

512

Primer	5'→3' sequence	Tm	Position	Amplicon size
Gene expression experiment				
qGBP1_F	TGGCTGAGAAGATGGAGAAG	58.1	E10	97 bp
qGBP1_R	TCCTGAATTAGTCGGGCTTG	60.2	E11	
qYWHAZ_F	TGATGATAAGAAAGGGATTGTGG	59.4	E3	134 bp
qYWHAZ_R	GTTCAGCAATGGCTTCATCA	61.3	E4	
qRPL32_F	CACCAGTCAGACCGATATGTCAA	61.1	E1	70 bp
qRPL32_R	CGCACCCTGTTGTCAATGC	61.1	E2	
Genotyping and allele-specific assay				
gGBP1_F	AGACCTAGAATCTCCACAGAATTTCCA	64.2	E11	105 bp
gGBP1_R	GGAAAGGACAGTTCGCTTCTCTAG	62.7	E11	
Probe allele A	VIC-CTGGGTGATAAATAAAT-NFQ		E11	
Probe allele G	FAM-TGGGTGATGAATAAAT-NFQ		E11	
3'RACE				
Oligo(dT)	ACTGGAAGAATTCTCGGCCGCAG $(T)_{30}$VN			
UAP	ACTGGAAGAATTCTCGGCCGCAG	70.3		
3'RACE_1	CTTCAGGAACAAGCCCGACT	62.8	E10-E11	1200-770 bp
3'RACE_2	GACCAGAAGACCCTGAGCAC	61.2	E11	750-350 bp
Polyadenylation usage ratio				
GBP1-fam	FAM-CTTCAGGAACAAGCCCGACTA	65.1	E10	
GBP1-SHORT_R	TTTTTTTTCGCTTCTCTAGCCCATT	64.8	E11	725 bp
GBP1-LONG_R	TCGAGCAGGAAAGGACAGTTC	62.9	E11	740 bp
Promoter sequencing				
prmGBP1_F	CCGGGATCTGGAGAGAACCT	63.9	Promoter	1029 bp
prmGBP1_R	TTCACGGGAGGGTTTGACTG	64.5	E1	
Exon 11 sequencing				
GBP1e11_F	GCCCGACTAATTCAGGAAGGA	63.6	E11	1078 bp
GBP1e11_R	AGGGCTTTTGACAACTGCAA	61.9	E11	

513

514

Supplementary Table S2. List of polymorphisms identified during the sequencing of the *GBP1* promoter and 5' and 3'UTR. Positions in promoter and 5'UTR are given in relation to the ATG START codon. Positions in the 3'UTR are counted from the TAA STOP codon.

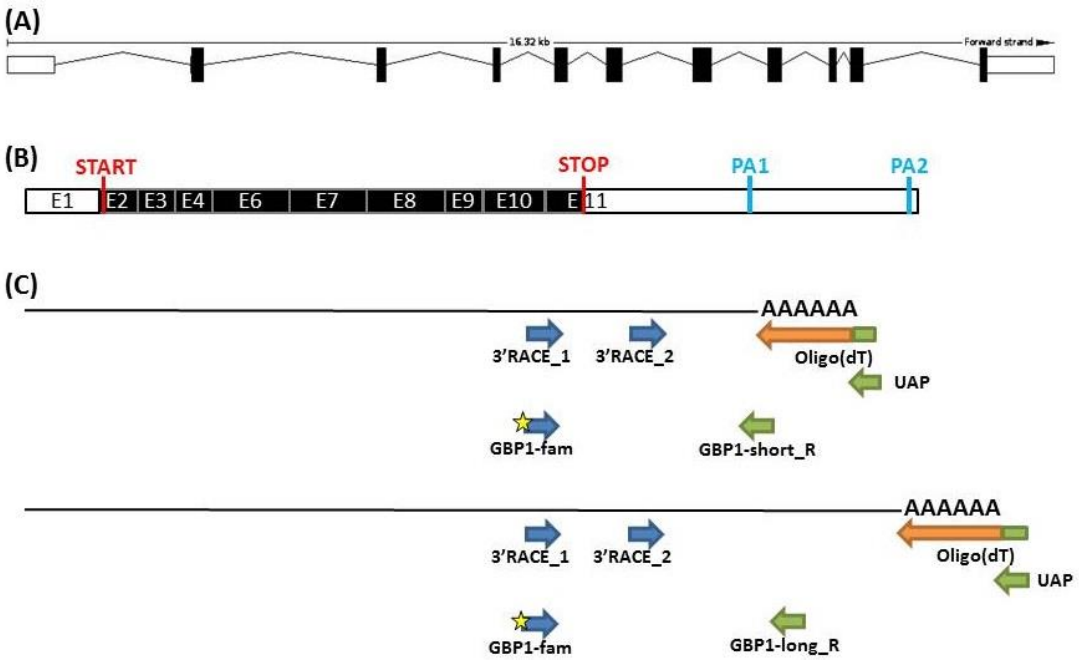
Polymorphism	Change	Position	WUR100125 AA pigs	WUR1000125 GG pigs
5'UTR				
rs335275118	G>T	-748	G	T
novel	G>C	-651	G	G/C
novel	C>A	-608	C	C/A
novel	T>C	-311	T	C
novel	T>C	-310	T/C	C
novel	INDEL	-307	delC	C
3'UTR				
rs339886073	G>T	*445	G	T
rs80863339	G>A	*540	A	G
rs342010514	C>T	*544	T	C
novel	A>G	*545	G	A
rs80800372**	A>G	*580	A	G
rs324386096	T>C	*610	T	C
novel	G>C	*718	C	G/C
novel	G>T	*742	G	G/T
novel	A>T	*743	A	A/T
rs323595619	T>C	*859	T	C

**rs80800372 = WUR1000125

Supplementary Table S3. *In silico* analysis of putative transcription factor-binding sites in the proximal promoter and exon 1 of *GBP1* gene, using PROMO and JASPAR tools. Sites potentially affected by polymorphisms found in the promoter region are highlighted.

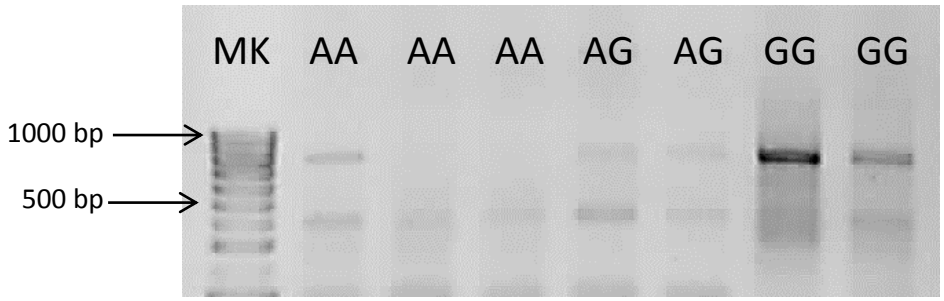
Supplementary Table S4. *In silico* analysis of the *GBP1* 3'UTR region for RNA-regulatory motifs, using the Reg-RNA tool.

Supplementary Figure S1. Location of primers used for the characterisation of *GBP1* 3'UTR and polyadenylation usage rate experiments. (A) Structure of *GBP1* in exons (boxes) and intron (lines) according to Ensembl (ENSSSCT00000007584). Coding sequences are filled in black and 5' and 3' regions in white. (B) Situation of START and STOP codons and proximal (PA1) and distal (PA2) polyadenylation signals in the mRNA. (C) Location of primers in relation to the short and long transcripts. Primer sequence and description is detailed in Supplementary Table S1.



Supplementary Figure S2. Characterisation of the 3'UTR region of the pig *GBP1* gene. (A) The 3'RACE experiment resulted in the amplification of two 3'UTR products in AA, AG and GG animals. MK – 100 bp ladder (HyperLadder, Bioline) (B) PCR fragments were sequenced and aligned. The two transcripts differed in the length of the 3'UTR by the alternative use of two polyadenylation signals (in pink). Ten SNP polymorphisms (described in SNPdb, in blue; novel, in red) were also identified.

(A)



(B)

```
Long-GG         GGACTGAGTCCGGCACGAGTTTGGTCAGGCCCACAACACATGCGAACATCGGGGACAACC 60
Long-AA         GGACTGAGTCCGGCACGAGTTTGGTCAGGCCCACAACACATGCGAACATCGGGGACAACC 60
Short-GG        GGACTGAGTCCGGCACGAGTTTGGTCAGGCCCACAACACATGCGAACATCGGGGACAACC 59
Short-AA        GGACTGAGTCCGGCACGAGTTTGGTCAGGCCCACAACACATGCGAACATCGGGGACAACC 59
                ************************************************************

Long-GG         CTTCTAGCTTCTCAGTGGAGGCAGCATCGTCTTCCACTGTTTCAAAACTGCCCAGTCTGT 120
Long-AA         CTTCTAGCTTCTCAGTGGAGGCAGCATCGTCTTCCACTGTTTCAAAACTGCCCAGTCTGT 120
Short-GG        CTTCTAGCTTCTCAGTGGAGGCAGCATCGTCTTCCACTGTTTCAAAACTGCCCAGTCTGT 119
Short-AA        CTTCTAGCTTCTCAGTGGAGGCAGCATCGTCTTCCACTGTTTCAAAACTGCCCAGTCTGT 119
                ************************************************************


Long-GG         CAGCAGTGATGGGATCCC GGGCACTGGGAATGCGCGATGCTTACTGGTGAATTCGAAGC 180
Long-AA         CAGCAGTGATGGGATCCC GGGCACTGGGAATGCGCGATGCTTACTGGTGAATTCGAAGC 180
Short-GG        CAGCAGTGATGGGATCCC GGGCACTGGGAATGCGCGATGCTTACTGGTGAATTCGAAGC 179
Short-AA        CAGCAGTGATGGGATCCC GGGCACTGGGAATGCGCGATGCTTACTGGTGAATTCGAAGC 179
                ****************** ****************************************


Long-GG         CAGAGCCAAGGCTCTTCTGGAGAGACCTAGAATCTCCACAGAATTTCCACAGG AGA  G 240
Long-AA         CAGAGCCAAGGCTCTTCTGGAGAGACCTAGAATCTCCACAGAATTTCCACAGG AGA  G 240
Short-GG        CAGAGCCAAGGCTCTTCTGGAGAGACCTAGAATCTCCACAGAATTTCCACAGG AGA  G 239
Short-AA        CAGAGCCAAGGCTCTTCTGGAGAGACCTAGAATCTCCACAGAATTTCCACAGG AGA  G 239
                ************************************************** *** .*


Long-GG         AAAAACCATTAGCATGACTGGCAGCTGGGTGAT       TGGGCTAGAGAAGCGAACTG 300
Long-AA         AAAAACCATTAGCATGACTGGCAGCTGGGTGAT       TGGGCTAGAGAAGCGAACTG 300
Short-GG        AAAAACCATTAGCATGACTGGCAGCTGGGTGAT       TGGGCTAGAGAAGCGAA--- 296
Short-AA        AAAAACCATTAGCATGACTGGCAGCTGGGTGAT       TGGGCTAGAGAAGCGAA--- 296
                ******************************** ***********************


Long-GG         TCC TTCCTGCTCGATTCGCGCAGATTCTAACTTTACTAGGTGGGACTCTCTGGAATTTT 360
Long-AA         TCC TTCCTGCTCGATTCGCGCAGATTCTAACTTTACTAGGTGGGACTCTCTGGAATTTT 360
Short-GG        ------------------------------------------------------------
Short-AA        ------------------------------------------------------------
```

```
Long-GG         AGGTTACAGTGGACTACACAGTGACCTGAAAACAGTTTCCCATGGCGTTTG GGCAATTT 420
Long-AA         AGGTTACAGTGGACTACACAGTGACCTGAAAACAGTTTCCCATGGCGTTTG GGCAATTT 420
Short-GG        ------------------------------------------------------------
Short-AA        ------------------------------------------------------------

Long-GG         ACAGTCTGCAAAGAA  ATGTGAAATGACAACAGAAACTGTGTTCGAAAACTGAGCTAAC 480
Long-AA         ACAGTCTGCAAAGAA  ATGTGAAATGACAACAGAAACTGTGTTCGAAAACTGAGCTAAC 480
Short-GG        ------------------------------------------------------------
Short-AA        ------------------------------------------------------------

Long-GG         TTAAGCGGCTAGACGGTTTAACCCTAGAGTTTAAGCTATCTTTTCCAAATTCTTCGCCAT 540
Long-AA         TTAAGCGGCTAGACGGTTTAACCCTAGAGTTTAAGCTATCTTTTCCAAATTCTTCGCCAT 540
Short-GG        ------------------------------------------------------------
Short-AA        ------------------------------------------------------------

Long-GG         CATACATAAAAA TTATTTTTGCCCTAGAGAATATGAATTGCTTTTGACATTTTTGCCCA 600
Long-AA         CATACATAAAAA TTATTTTTGCCCTAGAGAATATGAATTGCTTTTGACATTTTTGCCCA 600
Short-GG        ------------------------------------------------------------
Short-AA        ------------------------------------------------------------

Long-GG         GTTAAATAATGCTCTTGCTATTACTTAGTATATAGACTTTATTGCAGTTGTCAAAAGCCC 660
Long-AA         GTTAAATAATGCTCTTGCTATTACTTAGTATATAGACTTTATTGCAGTTGTCAAAAGCCC 660
Short-GG        ------------------------------------------------------------
Short-AA        ------------------------------------------------------------

Long-GG         TAGGTAAATGGGAAGACGATTAAGAGTATTTTCGAGCTGGA      CTGTGCTTCACTG 720
Long-AA         TAGGTAAATGGGAAGACGATTAAGAGTATTTTCGAGCTGGA      CTGTGCTTCACTG 720
Short-GG        ------------------------------------------------------------
Short-AA        ------------------------------------------------------------

Long-GG         AGCAAAAAAAAAAAAAAAAAAAA...CTGCGGCCGAGAATTCTTCCAGT
Long-AA         AGCAAAAAAAAAAAAAAAAAAAA...CTGCGGCCGAGAATTCTTCCAGT
Short-GG        ---AAAAAAAAAAAAAAAAAAAA...CTGCGGCCGAGAATTCTTCCAGT
Short-AA        ---AAAAAAAAAAAAAAAAAAAA...CTGCGGCCGAGAATTCTTCCAGT
```

Supplementary Figure S3. Characterisation of sequence variation in the promoter (in italics) and 5'UTR (exon 1) of the pig *GBP1* gene. Underlined: In pink, two putative overlapping sites for interferon regulatory factor (IRF-1) and NFκB; in red, a putative p53 binding element (human); in blue, a NFκB-binding motif; in green, ISRE (IFN-α stimulated response element). Polymorphic sites identified during the sequencing experiment are highlighter in black. Number on the right-hand side indicate bp from the ATG START codon in exon 2.

CCGGGATCTGGAGAGAACCTTGTAACCATCTCCCTTTCTCCTCACTCCCCTCTTCCTCGT -1034

CCAGGGCGAGAAAAAGCAGTGAGCTTAAGGGTAAACAGAGAATCAAATCTGTATCCACCT -974

CTGACGAGCTTGGTTGACAGGAACGGGCATCATCACCCACCCTCCTCATCATCACCCACC -914

CTCCACTGTAAGATGGAGACAGTCCCTCTTCTTTTGCCTGGTTCCGGTCAGGACTGAATT -854

GAACATTAACACAGGTAACTCTTAGAAAACAATCAAGACATAAAGCTGTGACTTCCCCTG -794

*CTCTTTGGCAGACACAGAAAGGCAGTACAGTGAAGTC*AGAAGGGT TGAGATCAAGGATG -734

ATTTGGTTCAGATTATTTACAGGGCTGCGATGTGGGACAAGAGGGTACATTTTCTGTGCC -674

TTAGTTTCTACAAAACG**ATGGG** **AAATGTT**TGTTCTAAGTTATTGAACGCATGTAGAGCG -614

GTCAG ACAGAGCCTGGAACCTGAGCACAGGACATGCAGCCCCTTCCTGCGTGTGGGAAT -554

TAAGTTCCAGGGACTGT**GTGTTCATGAAAGCGCCAGATCTGTCC**TCTTCTTCACTGACCC -494

CACACATAAAGGAGGAGCCTGGTTCACTGTGAGGCCGTTTTTAAGGAAATTAAACTTAAA -434

GATGAGGAGGCTTCCTCATTCTAAATAGTTTTTCAAACAGACCCCAATCCTGAGATATAG -374

TCCATTATATTTAAACATCTAGTAGACATGTTTTAAAACAGCAAATGATCTCTAAGATCT -314

CT CC CCCCCCCAAAAAAGAACGTGAAGATCATGCCAAATCCATTTACCTTCCTCCCA -254

GGATGGCTTTT**AGAAATTCCT**TTTATGGTTGTTGAGTCATTGCTTTGTATTCATT**GCTTT** -194

CAGTTTCATATTTATTCTAAGTCTATTACAGAGGTTGCTTTGCTTCTGACTCGGCTCTAG -134

AGGGAATCAGTAAAGCTCCTCGACACTGGCTGTGTGGACTAACAGTCAAACCCTCCCGTG -74

AAACAGAGAAGTTACAGAGAAGTCCACTCGTCTCACTGAGAAGAGGAAAGAACTCTCAAT -14

GAG -11 <[INTRON 1]

Supplementary Figure S4. Regulatory, structural and sequence variation analysis of the pig *GBP1* 3'UTR region. The sequence from the STOP codon (boxed in grey) in exon 11 is indicated. Four interferon-response elements are underlined. The 10 SNP polymorphisms identified during the sequencing of the 3'UTR region are highlighted in black except the WUR1000125 mutation, which is highlighted in green. Numbers on the right-hand side indicate nucleotide positions from the TAA STOP codon.

EXON11<TAAAGAACTGGAGAAGAGCGCTTTCCCGGTCCCGCTTAGCCATGGTCTTGCTCA *51
AGTAGTTTAGAATTAAGGAAAATGTTGCCAAACCTGATGATAATTACATTTCACATTGGT *111
ATTA**CACAAAGAACTCGCACATCACGCAGC**AGGGTACCTGAAATCATCTCGACCTTCCTC *171
ACCACACCAAAGGGGGGACAGGATACGCATTTCACCTCTGCACCCGCCCAGATGGCACCA *231
CGGTCTGGTTCCAATCAGGAGCTTCCTCTTCCAGATGACCGCCAGCTAGACCAGAAGACC *291
CTGAGCACCGTCTCGGGACTGAGTCCGGCACGAGTTTGGTCAGGCCCACAACACATGCGA *351
ACATCGGGGACAACCCTTCTAGCTTCTCAGTGGAGGCAGCATCGTCTTCCACTGTTTCAA *411
AACTGCCCAGTCTGTCAGCAGTGATGGGATCCC GGGCACTGGGAATGCGCGATGCTTAC *471
TGGTGAATTCGAAGCCAGAGCCAAGGCTCTTCTGGAGAGACCTA**GAATCTCCACAGAATT** *531
TCCACAGG AGA **GAAAAACCATTAGCATGA**CTGGCAGCTGGGTGAT**A**AATAAATGGGC *591
TAGAGAAGCGAACTGTCC TTCCTGCTCGATTCGCGCAGATTCTAACTTTACTAGGTGGG *651
ACTCTCTGGAATTTTAGGTTACAGTGGACTACACAGTGACCTGAAAACAGTTTCCCATGG *711
CGTTTG GGCAATTTACAGTCTGCAAAGAA ATGTGAAATGACAACAGAAACTGTGTTC *771
GAAAACTGAGCTAACTTAAGCGGCTAGACGGTTTAACCCTAGAGTTTAAGCTATCTTTTC *831
CAAATTCTTCGCCATCATACATAAAAA TTATTTTTGCCCTAGAGAATATG**AATTGCTTT** *891
TGACATTTTTGCCCAGTTAAATAATGCTCTTGCTATTACTTAGTATATAGACTTTATTGC *951
AGTTGTCAAAAGCCCTAGGTAAATGGGAAGACGATTAAGAGTATTTTCGAGCTGGAAATA *1011
AACTGTGCTTCACTGAGCTTTA *1071